DOI: 10.15960/j.cnki.issn.1007-6093.2017.04.007

矩阵优化扰动性分析的若干进展*

丁 超[1,†]

摘要 由于近年来实际问题特别是大数据应用的发展, 矩阵优化问题越来越得到优化研究者, 甚至是其他领域的研究者的高度关注, 成为热点问题. 优化问题的扰动性分析是优化理论研究的基础与核心, 为包括算法设计在内的优化研究提供重要的理论基础. 由于矩阵优化问题的非多面体性, 使得相应扰动分析理论的研究本质上与经典的多面体优化问题(非线性规划)不同. 结合文献 [1,2], 简要介绍矩阵优化扰动性分析方面取得的若干最新进展.

关键词 矩阵优化, 扰动性分析, 鲁棒孤立平稳性, 平稳性, 度量次正则
中图分类号 O221
2010 数学分类号 65K05, 90C25, 90C31

Preemptive online algorithms for scheduling*

DING Chao[1,†]

Abstract Matrix optimization problems (MOPs) have been recognized in recent years to be a powerful tool to model many important applications arising from emerging fields such as data science within and beyond the optimization community. Perturbation analysis of optimization problems play a fundamental and crucial role in optimization, which provided important theoretical foundation for algorithm designing and others. Science MOPs are non-polyhedral, the corresponding analysis is totally different from that of the classical polyhedral case (e.g., the nonlinear programming). Basing on results obtained in [1,2], we summary the recent progress on perturbation analysis of MOPs.

Keywords matrix optimization, perturbation analysis, robustly isolated calmness, calmness, metric subregularity
Chinese Library Classification O221
2010 Mathematics Subject Classification 65K05, 90C25, 90C31

0 引 言

矩阵优化问题泛指一类自变量为矩阵的优化问题, 其一般形式可以表示为以下抽象形式

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x)+\theta(q(x)) \\ \text{s.t.} \quad & l(x)\in\mathcal{Q}, \end{aligned} \tag{0.1}$$

收稿日期: 2017-08-15
* 基金项目: 国家自然科学基金(Nos. 11671387, 11301515)
1.中国科学院数学与系统科学研究院, 北京 100190; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
† 通信作者 E-mail: dingchao@amss.ac.cn

其中 $\mathbb{X}$ 是一般有限维欧氏空间, $f:\mathbb{X}\to(-\infty,+\infty)$ 为给定的光滑函数(例如, data-fitting 项), 而 $\theta:\mathbb{V}\to(-\infty,+\infty]$ 为定义在矩阵空间 $\mathbb{V}\equiv\mathbb{R}^{n\times m}$ $(n\leqslant m)$ 或者对称矩阵空间 $\mathcal{S}^n$ 上的一般非光滑真 (proper) 闭凸函数(例如, 特定数据集合的指示函数、各种数据正则项等), $q:\mathbb{X}\to\mathbb{V}$ 和 $l:\mathbb{X}\to\mathbb{R}^e$ 为给定的两个光滑函数, $\mathcal{Q}\subseteq\mathbb{R}^e$ 是一给定的凸多面体集(例如, $\{0\}\subseteq\mathbb{R}^e$、$\mathbb{R}^e_+$、带上下界的盒子约束).

矩阵优化 (0.1) 作为一个一般化的抽象优化框架, 包含了许多重要模型. 当矩阵空间 $\mathbb{V}=\mathbb{R}^{n\times m}$ 退化成向量空间 $\mathbb{R}^{n\times 1}$ 时, 模型 (0.1) 包含了经典的非线性规划问题 (NLP). 除了经典问题, 矩阵优化 (0.1) 当然也包含了许多新的应用. 例如, 当 $\mathbb{V}\equiv\mathcal{S}^n$, $\theta\equiv\delta_{\mathcal{S}^n_+}$, 其中 $\delta_{\mathcal{S}^n_+}$ 为半正定矩阵锥 $\mathcal{S}^n_+$ 的指示函数, 模型 (0.1) 就化为一般的非线性半正定规划 (NLSDP). 进一步, 当 $\mathbb{X}\equiv\mathcal{S}^n$, $q(x)\equiv x$ 并且 f 与 l 为线性函数时, (0.1) 就化为一般线性半正定规划 (SDP) 问题. 非线性半正定优化 (NLSDP) 和线性半正定规划 (SDP) 在诸如金融工程、组合图论、信号处理与控制等领域有着广泛的应用, 具体实例参见 [3-5]. 除了对称矩阵的情形, 模型 (0.1) 还包含了许多在诸如统计、数据挖掘、机器学习、信息学等领域重要的非对称矩阵相关应用. 例如, 当 $\mathbb{X}=\mathbb{V}\equiv\mathbb{R}^{m\times n}$, $q(x)\equiv x$, 而函数 θ 可以是矩阵完成问题[6-8]中的核范数 $\|\cdot\|_*$(矩阵的奇异值的和); 矩阵逼近问题[9-11]中的谱范数 $\|\cdot\|_2$(矩阵的最大奇异值); 或者是快速混合马氏链问题[12-13]中的 Ky Fan k-范数 $\|\cdot\|_{(k)}$(矩阵的最大 k 个奇异值之和). 基于这些重要应用, 矩阵优化模型 (0.1), 近年来受到广泛的关注, 在理论、算法与实际应用方面都取得了许多重要进展.

这里, 我们将针对矩阵优化问题 (0.1), 结合一些最新成果[1-2], 简要介绍一下相关扰动性分析的若干进展. 优化问题的扰动分析就是研究解(最优解或者稳定点)随扰动变量改变的变化规律, 从而对优化问题求解算法设计在内的其他方向提供重要的理论指导. 优化问题的扰动分析研究一直是优化问题理论研究的核心, 哪怕是针对其中最重要的结果, 做一个简单综述都远远超越了作者能力范围. 因此, 关于这一方面, 建议读者参考以下经典文献以及其中提到的参考文献[14-20]. 这里, 我们想强调的是当优化问题是多面体时, 例如模型 (0.1) 中的非光滑函数 θ 是凸多面体函数(其上图为凸多面体), 相应的扰动性分析通过过去三十年的研究已经相当完备了[20,23]. 然而, 随着实际应用的发展, 特别是近年来, 诸如统计优化在内的大数据科学相关应用的不断涌现, 人们发现要求解的优化问题往往是非多面体的. 为了更好的理解优化问题的理论本质, 设计求解适合大数据相关的优化问题特点的有效的算法, 研究非多面体的矩阵优化问题 (0.1) 的扰动分析就显得十分重要.

文献 [1-2] 分别研究了矩阵优化问题 (0.1) 解集的两个不同的扰动分析性质: 鲁棒孤立平稳性和平稳性. 令 $\mathbb{E}$ 和 $\mathbb{F}$ 为两个有限维空间. 假设 $\Upsilon:\mathbb{E}\rightrightarrows\mathbb{F}$ 为一给定的集值映射, 并且 $(\overline{p},\overline{q})\in\operatorname{gph}\Upsilon$, 即 $\overline{q}\in\Upsilon(\overline{p})$, 那么相应的平稳性和鲁棒孤立平稳性的定义分别如下给出:

定义 0.1(平稳性) 集值映射 $\Upsilon:\mathbb{E}\rightrightarrows\mathbb{F}$ 在 $\overline{p}$ 关于 $\overline{q}\in\Upsilon(\overline{p})$ 是平稳的, 如果存在常数 $\kappa>0$, $\overline{p}$ 的邻域 $\mathcal{U}$ 以及 $\overline{q}$ 的邻域 $\mathcal{V}$ 使得

$$\Upsilon(p)\cap\mathcal{V}\subset\Upsilon(\overline{p})+\kappa\|p-\overline{p}\|\mathbb{B},\quad\forall\, p\in\mathcal{U},\tag{0.2}$$

其中 $\mathbb{B}\subseteq\mathbb{F}$ 为空间 $\mathbb{F}$ 里的单位球.

定义 0.2(鲁棒孤立平稳性) 集值映射 $\Upsilon:\mathbb{E}\rightrightarrows\mathbb{F}$ 在 $\overline{p}$ 关于 $\overline{q}\in\Upsilon(\overline{p})$ 是孤立平稳的, 如果存在常数 $\kappa>0$, $\overline{p}$ 的邻域 $\mathcal{U}$ 以及$\overline{q}$ 的邻域 $\mathcal{V}$ 使得

$$\Upsilon(p)\cap\mathcal{V}\subset\{\overline{q}\}+\kappa\|p-\overline{p}\|\mathbb{B},\quad\forall\, p\in\mathcal{U},\tag{0.3}$$

其中 $\mathbb{B} \subseteq \mathbb{F}$ 为空间 F 里的单位球. 更进一步, 我们说 Υ 在 $\overline{p}$ 关于 $\overline{q}$ 是鲁棒孤立平稳的, 如果 (0.3) 成立并且对于任给的 $p \in \mathcal{U}$, $\Upsilon(p) \cap \mathcal{V} \neq \varnothing$.

对于一般(非凸)含非多面体约束矩阵优化问题, 文献 [1] 刻画了其 KKT 系统解映射在最优解处的鲁棒孤立平稳性, 证明了这一性质等价于二阶最优性条件和所谓严格 Robinson 约束品性同时成立(详见节 1 的定理 1.1). 从上述定义 0.2 和定义 0.1 不难看出, 鲁棒孤立平稳性意味着非扰动问题的最优解在一个邻域内是唯一的, 然而对于一些实际应用最优解往往不唯一, 即鲁棒孤立平稳性对于一些实际应用可能不成立, 这时我们需要研究相应的平稳性. 针对一般凸矩阵优化问题(问题 (0.1) 的特例), 文献 [2] 给出了扰动问题的最优解集合映射在某一最优解处平稳性成立充分性条件(详见节 2 的定理 2.1). 如前所述, 扰动性分析是优化理论研究的核心, 为诸如算法设计等提供重要理论. 这里, 通过对凸矩阵优化问题增广拉格朗日法的收敛性分析, 来展示相应的扰动分析结果不仅具有自身理论研究价值, 而且对解决实际问题有十分重要的指导意义, 是设计求解适应大数据背景的实际应用的快速算法的关键(详见节 3).

以下是一些常用的基本定义与符号:

(1) 如果 $\Omega \subseteq \mathbb{E}$ 为有限维线性空间 $\mathbb{E}$ 中的子集, 对于任给的 $w \in \mathbb{E}$, 令 $\mathrm{dist}(w, \Omega) := \inf\{\|u - w\| \mid u \in \Omega\}$.

(2) 如果 $\Omega \subseteq \mathbb{E}$ 为有限维线性空间 $\mathbb{E}$ 中的锥, 我们用 Ω^* 表示相应的对偶锥, 即

$$\Omega^* := \{w \in \mathbb{E} \mid \langle u, w\rangle \geqslant 0, \quad \forall u \in \Omega\}.$$

(3) 如果 $\Omega \subseteq \mathbb{E}$ 为有限维线性空间 E 中的子集, 我们用 $\mathcal{T}_\Omega(u)$ 表示 Ω 在 $u \in \Omega$ 的切锥[20], 即

$$\mathcal{T}_\Omega(u) := \{d \in \mathcal{E} \mid \exists u^k \to u,\ u^k \in \Omega \text{ 并且 } t^k \downarrow 0 \text{ 使得 } (u^k - u)/t^k \to d\}.$$

(4) 如果 $\Omega \subseteq \mathbb{E}$ 为有限维线性空间 E 中的闭集, 我们分别用 $\mathcal{T}^{i,2}_\Omega(u, d)$ 和 $\mathcal{T}^2_\Omega(u, d)$ 表示集合 Ω 在 $u \in \Omega$ 处沿方向 $d \in \mathbb{E}$ 的内/外二阶切集[15(3.49)和(3.50)], 即

$$\mathcal{T}^{i,2}_\Omega(u, d) := \Big\{w \in \mathbb{E} \mid \mathrm{dist}\Big(u + td + \frac{1}{2}t^2 w, \Omega\Big) = o(t^2), t \geqslant 0\Big\}$$

和

$$\mathcal{T}^2_\Omega(u, d) := \Big\{w \in \mathbb{E} \mid \exists\, t_k \downarrow 0 \text{ 使得 } \mathrm{dist}\Big(u + t_k d + \frac{1}{2}t_k^2 w, \Omega\Big) = o(t_k^2)\Big\}.$$

(5) 如果 $\Omega \subseteq \mathbb{E}$ 为有限维线性空间 $\mathbb{E}$ 中的子集, 我们用 $\sigma(\cdot, \Omega)$ 表示 Ω 的支撑函数, 即

$$\sigma(w, \Omega) := \sup\{\langle u, w\rangle \mid u \in \Omega\}, \quad w \in \mathbb{E}.$$

1 鲁棒孤立平稳性的刻画

为简单起见, 我们在本节假设矩阵优化模型 (0.1) 中的非光滑函数 θ 为空间 V 中的给定非空闭凸集合 Σ 的指示函数 δ_Σ. 我们可以将 (0.1) 表示为如下抽象优化问题,

$$\begin{cases} \min & f(x) \\ \text{s.t.} & G(x) \in \mathcal{K}, \end{cases} \tag{1.1}$$

其中 $\mathbb{Y}:=\mathbb{V}\times\mathbb{R}^e$, $G(x):=(q(x),l(x)):\mathbb{X}\to\mathbb{Y}$ 并且 $\mathcal{K}:=\Sigma\times\mathcal{Q}\subseteq\mathbb{Y}$. 对于问题 (1.1), 考虑以下线性扰动问题,

$$\begin{array}{ll}\min & f(x)-\langle a,x\rangle\\ \text{s.t.} & G(x)+b\in\mathcal{K},\end{array}\tag{1.2}$$

$(a,b)\in\mathbb{X}\times\mathbb{Y}$ 为扰动参数. 对于给定的 $(a,b)\in\mathbb{X}\times\mathbb{Y}$, 我们用 $X(a,b)$ 表示问题 (1.2) 所有的局部最优解. 如果存在 x 的一个开邻域 $\mathcal{V}$ 使得 $X(a,b)\cap\mathcal{V}=\{x\}$, 我们说 $x\in X(a,b)$ 是孤立的. 令 $\Phi(a,b)$ 为问题 (1.2) 的所有可行解构成的集合, 即

$$\Phi(a,b):=\{x\in\mathbb{X}\mid G(x)+b\in\mathcal{K}\},\quad (a,b)\in\mathbb{X}\times\mathbb{Y}.\tag{1.3}$$

令 $L:\mathbb{X}\times\mathbb{Y}\to\mathbb{R}$ 为问题 (1.2) 在 $(a,b)=(0,0)$ 时的拉格朗日函数, 即

$$L(x;y):=f(x)+\langle y,G(x)\rangle,\quad (x,y)\in\mathbb{X}\times\mathbb{Y}.\tag{1.4}$$

对于给定的 $y\in\mathbb{Y}$, 我们用 $L'_x(x;y)$ 来表示 $L(\cdot;y)$ 在 $x\in X$ 的导数并且用 $\nabla_x L(x;y)$ 表示 $L'_x(x;y)$ 的伴随算子. 对于给定的扰动参数 (a,b), 问题 (1.2) 的 Karush-Kuhn-Tucker (KKT) 最优性条件可以表示为

$$\begin{cases}a=\nabla_x L(x;y),\\ b\in -G(x)+\partial\sigma(y,\mathcal{K})\end{cases}\quad\Longleftrightarrow\quad\begin{cases}a=\nabla_x L(x;y),\\ y\in\mathcal{N}_{\mathcal{K}}(G(x)+b),\end{cases}\tag{1.5}$$

其中 $\partial\sigma(y,\mathcal{K})$ 是 $\mathcal{K}$ 的支撑函数 $\sigma(\cdot,\mathcal{K})$ 在 y 点处的次微分, $\mathcal{N}_{\mathcal{K}}(z)$ 为 $\mathcal{K}$ 在 $z\in\mathcal{K}$ 处的法锥(在凸分析的意义下见文[21]). 对于给定的 (a,b), 我们用 $S_{\rm KKT}(a,b)$ 表示 KKT 系统 (1.5) 的解集合, 用 $X_{\rm KKT}(a,b)$ 表示问题 (1.2) 的稳定点集合, 即 $X_{\rm KKT}(a,b):=\{x\in\mathbb{X}\mid$ 存在 $y\in\mathbb{Y}$ 使得 (1.5) 在 (x,y) 成立$\}$. 定义相应的拉格朗日乘子集合为 $M(x,a,b):=\{y\in\mathbb{Y}\mid(x,y)\in S_{\rm KKT}(a,b)\}$. 对于 $(a,b)=(0,0)$, 我们说问题 (1.2) 对应的 Robinson 约束品性 (RCQ) 在可行点 $x\in\Phi(0,0)$ 成立, 如果以下条件成立

$$G'(x)\mathbb{X}+\mathcal{T}_{\mathcal{K}}(G(x))=\mathbb{Y}.\tag{1.6}$$

特别地, RCQ 在一个局部极小点 $x\in X(0,0)$ 成立当且仅当 $M(x,0,0)$ 是 $\mathbb{Y}$ 中的一个非空凸的紧子集[22].

显然, 如果 $S_{\rm KKT}$ 在零点关于 $\overline{\beta}=(\overline{x},\overline{y})\in S_{\rm KKT}(0,0)$ 是孤立平稳的, 那么我们知道 $S_{\rm KKT}(0,0)$ 是单点集, 即 $S_{\rm KKT}(0,0)=\{(\overline{x},\overline{y})\}$. 孤立平稳性是一个重要的扰动分析性质, 在不同的经典文献中(例如, 文 [16,23,24]) 也被称为局部上 Lipschitz 连续性并广泛研究. 一般而言, 孤立平稳性并不意味着鲁棒孤立平稳性(详见文 [25, Example 6.4]).

在进一步介绍问题 (1.2) 的(鲁棒)孤立平稳性的完整刻画之前, 我们首先简要回顾一下相关研究进展. 现有的研究主要集中在当集合 $\mathcal{K}$ 为多面体这一特殊情形下(即问题 (1.1) 中的非空闭凸集 $\mathcal{C}$ 为一给定的多面体). 换言之, 现有文献主要研究非线性规划稳定点映射的(鲁棒)孤立平稳性的刻画. 具体而言, 当问题 (1.2) 中的集合 $\mathcal{K}$ 为多面体的时候, 相应的(鲁棒)孤立平稳性的研究是相当完备的. 甚至在更一般的扰动框架下, 即问题 (1.2) 中的函数 f 和 G 满足以下的参数化形式: $f(x,c)$ 和 $G(x,c)$, $c\in E$ 其中 E 为一给定有限维线性空间, 人们得到了相关性质的完备数学刻画[23,26,27]. 例如, Dontchev 和

Rockafellar[23] 证明了对于参数化的扰动非线性规划问题的局部最优点, 稳定点映射 S_{KKT} 是鲁棒孤立平稳的当且仅当严格 Mangasarian-Fromovitz 约束品性 (strict MFCQ) 和二阶最优性条件 (SOSC) 同时成立. 那么以上关于多面体的结果能不能推广到非多面体的情形? 例如我们的矩阵优化模型 (0.1), 特别是半正定规划 (SDP) 的情况. 在回答这个问题之前, 让我们首先来看以下二次半正定凸优化例子

例 1.1 Bonnans 和 Shapiro[15,Example 4.54]:

$$\begin{array}{ll} \min & x_1 + x_1^2 + x_2^2 \\ \text{s.t.} & \mathrm{Diag}(x) + \varepsilon A \in \mathcal{S}_+^2, \end{array} \tag{1.7}$$

其中 $x = (x_1, x_2) \in \mathbb{R}^2$, $\mathrm{Diag}(x)$ 是一个 2×2 对角矩阵, 其第 i 个对角元素为 x_i, $i = 1, 2$, A 是 $\mathcal{S}^2$ 中的一个非对角矩阵, 并且 ε 为参数. 显然, (1.7) 的 Slater 条件成立. 当 $\varepsilon = 0$ 时, 优化问题 (1.7) 有唯一解 $\overline{x} = (0, 0)$ 以及与之对应的唯一的拉格朗日乘子 $\overline{Y} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$. 显然, 对于任给的参数 $\varepsilon \geqslant 0$, 问题 (1.7) 有唯一解 $X(\varepsilon) = (\overline{x}_1(\varepsilon), \overline{x}_2(\varepsilon))$ 其中当 $\varepsilon \to 0$ 时, $\overline{x}_2(\varepsilon)$ 与 $\varepsilon^{2/3}$ 同阶.

例 1.1 表明即使在强凸二次半正定优化问题的 Slater 条件成立(进而 MFCQ 成立)的前提下, 相应的 KKT 解映射 S_{KKT} 的平稳性不成立, 进而孤立平稳性也不成立. 这表明含非多面体的优化问题与多面体情形完全不同, 因为 Robinson[26] 证明了对多面体情形(非线性规划)而言, 如果所谓 MFCQ 和二阶最优性条件成立, 则 KKT 解映射 S_{KKT} 的平稳性必然成立. 与此同时, 例 1.1 也表明对于一般非多面体问题, 拉格朗日乘子唯一性与二阶最优性条件不能保证 KKT 解映射 S_{KKT} 的孤立平稳性. 同样这也与多面体情形即非线性规划问题完全不同, 因为我们知道根据文献 [28, Proposition 1.1], 对于非线性规划问题严格 Mangasarian-Fromovitz 约束品性 (MFCQ) 等价于拉格朗日乘子唯一性, 进而前述 Dontchev 和 Rockafellar 的关于多面体情形的经典结果也可等价的表述为 S_{KKT} 的鲁棒孤立平稳性等价于拉格朗日乘子唯一性和二阶最优性条件 (SOSC) 同时成立. 通过例 1.1 我们可以看到, 对于一般非多面体优化问题 (1.2) 相应的 KKT 解映射 S_{KKT} 的鲁棒孤立平稳性的刻画并不是非线性规划问题相关性质的简单推广, 其非多面体性在相应分析中起到了至关重要的作用.

下面我们利用相关集合投影算子微分性质这个简单的例子, 来具体说明非多面体集合与多面体集合在相关分析中的差异. 这些差异也使矩阵优化的扰动分析与经典非线性规划的扰动分析不同. 首先, 我们回顾所谓临界锥的定义. 假设 $\overline{x}$ 是问题 (1.2) 关于 $(a, b) = (0, 0)$ 的一个可行解. 问题 (1.2) 对于 $(a, b) = (0, 0)$ 在 $\overline{x}$ 的临界锥 $\mathcal{C}(\overline{x})$ 可以定义为

$$\mathcal{C}(\overline{x}) := \{d \in \mathbb{X} \mid G'(\overline{x})d \in \mathcal{T}_{\mathcal{K}}(G(\overline{x})), f'(\overline{x})d \leqslant 0\}. \tag{1.8}$$

特别地, 如果 $\overline{x}$ 是问题 (1.2) 对应 $(a, b) = (0, 0)$ 关于 $\overline{y} \in M(\overline{x}, 0, 0)$ 的稳定点, 则我们有

$$\begin{aligned} \mathcal{C}(\overline{x}) &= \{d \in X \mid G'(\overline{x})d \in \mathcal{T}_{\mathcal{K}}(G(\overline{x})), f'(\overline{x})d = 0\} \\ &= \{d \in \mathbb{X} \mid G'(\overline{x})d \in \mathcal{C}_{\mathcal{K}}(G(\overline{x}), \overline{y})\}, \end{aligned}$$

其中对于任给的 $A \in \mathcal{K}$, $\mathcal{C}_{\mathcal{K}}(A, B)$ 是集合 $\mathcal{K}$ 在 A 点关于 $B \in \mathcal{N}_{\mathcal{K}}(A)$ 的临界锥, 定义为

$$\mathcal{C}_{\mathcal{K}}(A, B) := \mathcal{T}_{\mathcal{K}}(A) \cap B^{\perp}, \tag{1.9}$$

并且对于任给的 $s \in \mathbb{Y}$, $s^{\perp} := \{z \in \mathbb{Y} \mid \langle z, s\rangle = 0\}$. 因为 $\mathcal{K}$ 是凸集, 所以对于任给的 $\overline{y} \in M(\overline{x}, 0, 0)$, 临界锥 $\mathcal{C}_{\mathcal{K}}(G(\overline{x}), \overline{y})$ 都是一个闭凸锥. 另一个重要概念是以下集合的 C^2-锥可约性.

定义 1.1[15,Definition 3.135] 闭凸集 $\mathcal{K}$ 被称为在 $\overline{A} \in \mathcal{K}$ 处 C^2-锥可约, 如果存在 $\overline{A}$ 的一个开邻域 $\mathcal{W} \subset Y$, 一个有限维线性空间 E 中的尖的闭凸锥 Δ(一个锥被称为是尖的当且仅当它的线性化空间为零点集 $\{0\}$)和一个二次连续可微的映射 $\Xi : \mathcal{W} \to \mathbb{E}$ 使得: (i) $\Xi(\overline{A}) = 0 \in \mathbb{E}$; (ii) 导数映射 $\Xi'(\overline{A}) : \mathbb{Y} \to \mathbb{E}$ 是满的; (iii) $\mathcal{K} \cap \mathcal{W} = \{A \in \mathcal{W} \mid \Xi(A) \in \Delta\}$. 如果 $\mathcal{K}$ 在任意点 $\overline{A} \in \mathcal{K}$ 都是 C^2-锥可约的, 则我们说 $\mathcal{K}$ 是 C^2-锥可约的.

一般而言, 即使对于闭凸集合, 相应的内/外二阶切集也不一定不相等[15,Section 3.3]. 但是, 根据文 [15, Proposition 3.136], 我们知道如果 $\mathcal{K}$ 是 C^2-锥可约的闭凸集时, 对任给的 $\overline{A} \in \mathcal{K}$ 和 $D \in \mathbb{Y}$, $\mathcal{T}^{i,2}_{\mathcal{K}}(\overline{A}, D) = \mathcal{T}^2_{\mathcal{K}}(\overline{A}, D)$ 成立. 此时, 我们称 $\mathcal{T}^2_{\mathcal{K}}(\overline{A}, D)$ 为 $\mathcal{K}$ 在 $\overline{A} \in \mathcal{K}$ 处沿方向 $D \in \mathbb{Y}$ 的二阶切集.

下面, 我们可以引进关于非空闭凸集 $\mathcal{K}$ 投影算子方向导数的结果. 我们用 $\Pi_{\mathcal{K}} : \mathbb{Y} \to \mathbb{Y}$ 表示 $\mathcal{K}$ 投影算子, 即对于任给的 $C \in \mathbb{Y}$,

$$\Pi_{\mathcal{K}}(C) := \operatorname{argmin}\Big\{\frac{1}{2}\|Y - C\|^2 \mid Y \in \mathcal{K}\Big\}.$$

假设 $\overline{B} \in \mathcal{N}_{\mathcal{K}}(\overline{A})$. 令 $C := \overline{A} + \overline{B}$. 如果 $\mathcal{K}$ 是多面体集[17,Theorem 4.1.1]投影算子是方向可微的并且其方向导数 $\Pi'_{\mathcal{K}}(C; \cdot)$ 有如下刻画, 即对于任给的方向 $H \in \mathbb{Y}$,

$$\Pi'_{\mathcal{K}}(C; H) = \operatorname{argmin}\{\|D - H\|^2 \mid D \in \mathcal{C}_{\mathcal{K}}(\overline{A}, \overline{B})\},$$

其中 $\mathcal{C}_{\mathcal{K}}(\overline{A}, \overline{B})$ 为集合 $\mathcal{K}$ 在 $\overline{A}$ 处关于 $\overline{B}$ 的临界锥 (1.9).

如果集合 $\mathcal{K}$ 是 C^2-锥可约凸集时, 根据文 [29, Theorem 7.2], 我们知道 $\Pi_{\mathcal{K}}$ 在 C 处依然是方向可微的并且其方向导数 $\Pi'_{\mathcal{K}}(C; H)$ 关于任给方向 $H \in \mathcal{Y}$ 可以如下刻画

$$\Pi'_{\mathcal{K}}(C; H) = \operatorname{argmin}\{\|D - H\|^2 - \sigma(\overline{B}, \mathcal{T}^2_{\mathcal{K}}(\overline{A}, D)) \mid D \in \mathcal{C}_{\mathcal{K}}(\overline{A}, \overline{B})\}, \tag{1.10}$$

其中 $\sigma(\cdot, \mathcal{T}^2_{\mathcal{K}}(\overline{A}, D))$ 为二阶切集 $\mathcal{T}^2_{\mathcal{K}}(\overline{A}, D)$ 的支撑函数. 该支撑函数被称为非多面体集合 $\mathcal{K}$ 的 σ-项. 它反映了非多面体的重要几何性质, 并在关于非多面体的优化问题扰动性分析中起到了至关重要的作用[15].

对于在 $(a, b) = (0, 0)$ 处问题 (1.2), 假设 $\overline{x}$ 为一可行解. 我们说严格 Robinson 约束品性在 $\overline{x}$ 关于 $\overline{y} \in M(\overline{x}, 0, 0) \neq \emptyset$ 成立, 如果

$$G'(\overline{x})\mathcal{X} + \mathcal{T}_{\mathcal{K}}(G(\overline{x})) \cap \overline{y}^{\perp} = \mathcal{Y}. \tag{1.11}$$

根据文 [15, Proposition 4.50], 如果严格 Robinson 约束品性成立, 则相应的拉格朗日乘子集合 $M(\overline{x}, 0, 0)$ 就为单点集.

根据相应方向导数 $\Pi'_{\mathcal{K}}(C; \cdot)$ 的刻画 (1.10) 及其他变分分析性质, 我们得到了以下关于稳定点映射 $S_{\rm KKT}$ 的鲁棒孤立平稳性的完整刻画[1]. 此外, 对于问题 (1.2) 且 $(a, b) = (0, 0)$, 假设相应的 Robinson 约束品性 (1.6) 在 $\overline{x}$ 成立, 则我们说相应的二阶充分性最优条件 (SOSC) 成立, 如果

$$\sup_{y \in M(\overline{x}, 0, 0)} \{\langle d, \nabla^2_{xx} L(\overline{x}; y) d\rangle - \sigma(y, \mathcal{T}^2_{\mathcal{K}}(G(\overline{x}), G'(\overline{x})d))\} > 0, \quad \forall d \in \mathcal{C}(\overline{x}) \setminus \{0\}. \tag{1.12}$$

定理 1.1 令 $\overline{x}$ 为问题 (1.2) 关于 $(a,b)=(0,0)$ 的可行解. 假设相应的 Robinson 约束品性 (1.6) 成立. 令 $\overline{y}\in M(\overline{x},0,0)\neq\varnothing$. 则以下结论等价:

(i) 严格 Robinson 约束品性 (1.11) 在 $\overline{x}$ 关于 $\overline{y}$ 成立并且问题 (1.2) 关于 $(a,b)=(0,0)$ 的二阶充分性最优性条件 (SOSC)(1.12) 在 $\overline{x}$ 处成立;

(ii) $\overline{x}$ 是问题 (1.2) 关于 $(a,b)=(0,0)$ 的一个局部最优解并且 $S_{\rm KKT}$ 在零点 $(a,b)=(0,0)$ 关于 $(\overline{x},\overline{y})$ 是鲁棒孤立平稳的;

(iii) $\overline{x}$ 是问题 (1.2) 关于 $(a,b)=(0,0)$ 的一个局部最优解并且 $S_{\rm KKT}$ 在零点 $(a,b)=(0,0)$ 关于 $(\overline{x},\overline{y})$ 是孤立平稳的.

2 最优解集平稳性的刻画

为了更好的讨论矩阵优化问题 (0.1) 另一个重要的扰动性质——最优解集平稳性, 在本节我们主要关注以下矩阵优化问题 (0.1) 的特例：

$$\begin{array}{ll}\min\limits_{X\in X} & \Phi(X):=h(\mathcal{F}X)+\langle C,X\rangle+\theta(X)\\ \text{s.t.} & \mathcal{A}X\in p+\mathcal{Q},\end{array}\tag{2.1}$$

其中 $\mathbb{X}\equiv\mathbb{V}$ 为 $n\times m$ 的矩阵空间 $\mathbb{R}^{n\times m}$ $(n\leqslant m)$ 或者对称矩阵空间 $\mathcal{S}^n$, 而 $C\in\mathbb{X}$ 和 $p\in\mathbb{R}^e$ 为给定数据, $\mathcal{F}:\mathbb{X}\to\mathbb{R}^d$ 和 $\mathcal{A}:\mathbb{X}\to\mathbb{R}^e$ 为给定线性映射, $h:\mathbb{R}^d\to(-\infty,+\infty]$ 在非空开凸的有效域 $\operatorname{dom}h$ 上连续可微并且 h 在 $\operatorname{dom}h$ 的任意凸紧子集上是强凸的, 函数 $\theta:\mathbb{X}\to(-\infty,+\infty]$ 为一真闭函数, 而 $\mathcal{Q}\subseteq\mathbb{R}^e$ 为一给定凸多面体锥. 问题 (2.1) 对应的对偶问题可以等价的表示为以下极小化形式

$$\begin{array}{ll}\min\limits_{(y,w,S)\in\mathbb{R}^e\times\mathbb{R}^d\times\mathbb{X}} & \Psi(y,w,S):=\delta_{\mathcal{Q}^*}(y)-\langle p,y\rangle+h^*(-w)+\theta^*(-S)\\ \text{s.t.} & \mathcal{A}^*y+\mathcal{F}^*w+S=C,\end{array}\tag{2.2}$$

其中 $\delta_{\mathcal{Q}^*}$ 为对偶锥 $\mathcal{Q}^*$ 的指示函数, $\mathcal{A}^*$, $\mathcal{F}^*$ 为 $\mathcal{A}$, $\mathcal{F}$ 相应的伴随算子, 而 h^*, θ^* 为 h, θ 的共轭函数.

对于原问题 (2.1), 我们考虑以下形式的扰动问题

$$\begin{array}{ll}\min\limits_{X\in\mathbb{X}} & \Phi(X)-\langle a,X\rangle\\ \text{s.t.} & \mathcal{A}X\in p+\mathcal{Q},\end{array}\tag{2.3}$$

其中$a\in\mathbb{X}$ 为扰动参数. 对于给定的 a, 我们用 $X(a)$ 表示问题 (2.1) 最优解集合. 令 $\overline{X}$ 为给定问题 (2.1) 的一个最优解, 即 $\overline{X}\in X(\overline{a})$ 其中 $\overline{a}=0$. 以下我们主要讨论集合映射 $X:\mathbb{X}\rightrightarrows\mathbb{X}$ 在 $\overline{X}$ 处关于扰动参数 $\overline{a}=0$ 的平稳性. 如引言中所述, 集值映射平稳性弱于我们前一节讨论的(鲁棒)孤立平稳性, 即如果集值映射是(鲁棒)孤立平稳的那么它也是平稳的.

为符号的简便, 记 $\mathbb{Z}:=\mathbb{R}^e\times\mathbb{R}^d\times\mathbb{X}$, 并且对于任给的 $y\in\mathbb{R}^e$, $w\in\mathbb{R}^d$ 以及 $S\in\mathbb{X}$, 令 $Z:=(y,w,S)$. 对偶问题 (2.2) 对应的 Lagrangian 函数 L 可以如下给出

$$L(Z;X):=\Psi(Z)+\langle X,\mathcal{A}^*y+\mathcal{F}^*w+S-C\rangle,\quad (Z,X)\in\mathbb{Z}\times\mathbb{X}.$$

令 $\phi:\mathbb{X}\to(-\infty,\infty]$ 为原问题 (2.1) 的本质目标函数, 即

$$\phi(X):=-\inf_{Z\in\mathbb{Z}}L(Z;X)=\begin{cases}\Phi(X), & \text{如果 } \mathcal{A}X\in p+\mathcal{Q},\\ +\infty, & \text{否则},\end{cases}\qquad \forall\, X\in\mathbb{X}. \tag{2.4}$$

利用本质目标函数 ϕ, 我们可以将原问题 (2.11) 可以表示成以下无约束问题

$$\min_{X\in\mathbb{X}}\quad \phi(X).$$

对于原问题 (2.1) 的本质目标函数 ϕ, 我们定义以下集值映射 $\mathcal{T}_\phi:\mathbb{X}\rightrightarrows\mathbb{X}$

$$\mathcal{T}_\phi(X):=\partial\phi(X),\quad X\in\mathbb{X}. \tag{2.5}$$

显然, 由于问题 (2.1) 是凸的, 集值映射 $\mathcal{T}_\phi$ 是极大单调的, 并且如果原问题 (2.1) 最优解集合 $X(0)$ 非空, 则

$$X(0)=\{X\in\mathbb{X}\mid 0\in\mathcal{T}_\phi(X)\}.$$

对于 $\mathcal{T}_\phi$ 以及任给的 $v\in\mathbb{X}$, 其图像和逆映射分别使用如下记号 $\operatorname{gph}\mathcal{T}_\phi:=\{(X,v)\in\mathbb{X}\times\mathbb{X}\mid v\in\mathcal{T}_\phi(X)\}$ 和 $\mathcal{T}_\phi^{-1}(v)=\{X\in\mathbb{X}\mid v\in\mathcal{T}_\phi(X)\}$. 显然, 对于任给的扰动参数 a, 我们有 $X(a)=\mathcal{T}_\phi^{-1}(a)$. 根据文 [16, Theorem 3H.3], 最优解集映射 X 的平稳性等价于以下 $\mathcal{T}_\phi$ 的度量次正则性[16,Section 3.8(3H)].

定义 2.1 集值映射 $\mathcal{T}_\phi:\mathbb{X}\rightrightarrows\mathbb{X}$ 在 $\overline{X}\in\mathcal{T}_\phi^{-1}(0)$ 关于零点是度量次正则的, 如果存在常数 $\kappa'>0$, $\overline{a}=0$ 的邻域 $\mathcal{U}$, 以及 $\overline{X}$ 的邻域 $\mathcal{V}$, 使得

$$\operatorname{dist}(X,\mathcal{T}_\phi^{-1}(0))\leqslant\kappa'\operatorname{dist}(0,\mathcal{T}_\phi(X)\cap\mathcal{U}),\quad \forall\, X\in\mathcal{V}.$$

众所周知, 如果函数 h 在 $\operatorname{dom}h$ 的任意凸紧子集上是强凸的, 则 $\mathcal{F}\overline{X}$ 在 $\overline{X}\in X(0)$ 是不变的(参见文献 [30] 或者 [31]). 因而, 对任意 $\overline{X}\in X(0)$ 下述变量是有定义的:

$$\overline{\xi}:=\mathcal{F}\overline{X},\quad \overline{\eta}:=\mathcal{F}^*\nabla h(\mathcal{F}\overline{X})+C. \tag{2.6}$$

我们定义集合 $\mathcal{V}_P:=\{X\in X\mid \mathcal{F}X=\overline{\xi}\}$ 以及两个映射 $\mathcal{G}_P^1,\mathcal{G}_P^2:\mathbb{R}^e\to\mathbb{X}$

$$\begin{array}{ll}\mathcal{G}_P^1(y):=(\partial\theta)^{-1}(\mathcal{A}^*y-\overline{\eta}), & \\ \mathcal{G}_P^2(y):=\{X\in\mathbb{X}\ \mid 0\in\mathcal{A}X-p+\mathcal{N}_{\mathcal{Q}^*}(y)\}, & \end{array}\qquad y\in\mathbb{R}^e.$$

根据问题 (2.1) 的 KKT 条件, 我们知道 $\overline{X}\in X(0)$, 则存在 Lagrangian 乘子 $\overline{y}\in\mathbb{R}^e$ 使得

$$\begin{cases}0\in C-\mathcal{A}^*\overline{y}+\mathcal{F}^*\nabla h(\mathcal{F}\overline{X})+\partial\theta(\overline{X}),\\ 0\in\mathcal{A}\overline{X}-p+\mathcal{N}_{\mathcal{Q}^*}(\overline{y}).\end{cases} \tag{2.7}$$

我们用 $\mathcal{M}_P(\overline{X})$ 来表示 $\overline{X}\in X(0)$ 所对应的所有 Lagrangian 乘子构成的集合. 显然, 我们有

$$X(0)=\mathcal{V}_P\cap\mathcal{G}_P^1(\overline{y})\cap\mathcal{G}_P^2(\overline{y}),\quad \forall\,\overline{y}\in\mathcal{M}_P(\overline{X}).$$

为了研究最优解集映射 X 的平稳性, 我们回顾集合的有界线性正则性[32].

定义 2.2 令 $D_1, D_2, \cdots, D_s \subseteq U$ 为 s 个一些闭凸集. 假设 $D := D_1 \cap D_2 \cap \cdots \cap D_s$ 非空, 则 $D_1, D_2, \cdots, D_s$ 被称为有界线性正则的, 如果对于每一个有界集 $\mathcal{B} \subseteq \mathbb{U}$, 存在常数 $\kappa > 0$ 使得

$$\mathrm{dist}(x, D) \leqslant \kappa \max\{\mathrm{dist}(x, D_1), \mathrm{dist}(x, D_2), \cdots, \mathrm{dist}(x, D_s)\}, \quad \forall\, x \in \mathcal{B}.$$

以上这个看似较为抽象的性质, 其实可以由一些常见的约束品性保证. 例如我们有以下结论[33,Corollary 3].

性质 2.1 令 $D_1, D_2, \cdots, D_s \subseteq \mathbb{U}$ 为 s 个闭凸集. 对于给定的 $s_0 \in \{0, 1, \cdots, s\}$, 假设 $D_1, D_2, \cdots, D_{s_0}$ 为 s_0 个多面体集. 那么, 保证 $D_1, D_2, \cdots, D_s$ 有界线性正则性成立的充分条件是

$$\bigcap_{i=1,2,\cdots,s_0} D_i \quad \bigcap \quad \bigcap_{i=s_0+1,\cdots,s} \mathrm{ri}\,(D_i) \neq \varnothing.$$

现在, 我们可以给出关于最优解集映射 X 的平稳性的一个充分性条件.

定理 2.1 令 $\overline{X}$ 为原问题 (2.1) 的一个最优解并且 $\mathcal{M}_P(\overline{X}) \neq \varnothing$. 假设以下三个条件成立

(a) 函数 h 在任意 $\mathrm{dom}\, h$ 的紧子集上是强凸的;

(b) 对于任给的 $\overline{v} \in \partial\theta(\overline{X})$, 映射 $\partial\theta$ 在 $\overline{X}$ 关于 $\overline{v}$ 是度量次正则的;

(c) 集合 $\mathcal{V}_P, \mathcal{G}_P^1(\overline{y}), \mathcal{G}_P^2(\overline{y})$ 在某一个 $\overline{y} \in \mathcal{M}_P(\overline{X})$ 满足有界线性正则性.

则最优解集映射 X 在 $\overline{X} \in X(0)$ 关于 $\overline{a} = 0$ 是平稳的.

在上述充分条件中, 条件 (b) 与 (c) 起到至关重要的作用. 自然人们会问上述条件何时成立? 特别是, 针对前述的重要矩阵优化问题, 这两个条件能不能成立?

首先, 我们考虑定理 2.1 中的条件 (b), 即 $\partial\theta$ 的度量次正则性. 对于大多数矩阵优化而言(例如, 半正定优化, 矩阵完成问题等), 问题 (2.1) 中的 θ 属于以下这类特殊的函数, 即矩阵的谱函数[34,35]. 它们可以被表达为以下两种形式之一:

$$\theta(X) = g(\sigma(X)), \quad X \in \mathbb{X}(\mathbb{R}^{n\times m} \text{ 其中 } n \leqslant m \text{ 或 } \mathbb{S}^n), \tag{2.8}$$

其中 $g : \mathbb{R}^n \to (-\infty, +\infty]$ 为真闭凸函数并且是绝对对称的, 或者

$$\theta(X) = g(\lambda(X)), \quad X \in \mathbb{S}^n, \tag{2.9}$$

其中 $g : \mathbb{R}^n \to (-\infty, +\infty]$ 为真闭凸函数并且是对称的.

性质 2.2 令 $g : \mathbb{R}^n \to (-\infty, +\infty]$ 为真闭凸的绝对对称函数. 令 $\theta : \mathbb{X} \to (-\infty, +\infty]$ 为 g 如 (2.8) 所定义的谱函数. 考虑任意给定的 $(\overline{X}, \overline{W}) \in \mathrm{gph}\,\partial\theta$. 假设次微分映射 ∂g 在 $\sigma(\overline{X})$ 处关于 $\sigma(\overline{W})$ 是度量次正则的, 则 $\partial\theta$ 在 $\overline{X}$ 处关于 $\overline{W}$ 是度量次正则的.

注 2.1 首先我们回顾一下, 如果一个集值映射是分片多面体的, 即它的图可以表示为有限多个多面体集合的并集. 在博士论文 [36] 中, 孙捷深入分析了分片线性-二次函数并证明了以下这个优美而又重要的结果: 一个真闭凸函数 p 是分片线性-二次的当且仅当 ∂p 是分片多面体的, 它也等价于其共轭函数 p^* 是分片线性-二次的(也可参见 [20, Theorem 11.14, Proposition 12.30]). 进一步, 通过利用 Robinson 的关于多面体映射的局部上 Lipschitz 连续性的一个经典结果[37], 我们得到对于任给的凸分片线性-二次函数 p

和任意点 $(x,w)\in \operatorname{gph}\partial p$, $(\partial p)^{-1}=\partial p^*$ 在 x 关于 w 是度量次正则的. 因而, 我们知道如果函数 g 是凸分片线性-二次的, 那么 (2.8) 所定义的谱函数 θ 在任给点 $X\in \operatorname{dom}\theta$ 是度量次正则的, 因而定理 2.1 中的条件 (b) 是成立的. 此类函数 g 的例子包含: 凸多面体的指示函数, 给定向量的前 k 个最大(绝对)值的和等. 由此, 性质 2.2 包含了许多关于核范数次微分以及半正定矩阵锥法锥的度量次正则性的经典结论, 如文[31, Proposition 11] 和 [38, Theorem 2.4](也参见 [39, Proposition 3.3]).

关于定理 2.1 中的条件 (c), 利用性质 2.1, 我们得到以下结果保证其充分性.

性质 2.3 如果以下任一条件满足, 那么集合 $\mathcal{V}_P, \mathcal{G}_P^1(\overline{y}), \mathcal{G}_P^2(\overline{y})$ 在某一个 $\overline{y}\in\mathcal{M}_P(\overline{X})$ 满足有界线性正则性:

(a) $\mathcal{G}_P^1(\overline{y})$ 是多面体集合;

(b) 存在 $\widehat{X}\in X(0)$ 使得 $\widehat{X}\in \operatorname{ri}(\mathcal{G}_P^1(\overline{y}))$.

在结束本节前, 我们针对性质 2.3 中的条件, 给出几个注释. 首先, 通过给出几个例子来说明条件 (a) 在实际应用中是可能成立的.

例 2.1 任给的 $X\in\mathbb{R}^n$, $\theta(X)=\|X\|_2$(即向量的 l_2-范数). 那么, $(\partial\theta)^{-1}(X)$ 对于任给的 $X\in\mathbb{R}^m$ 都是一个多面体集合[31,40], 从而条件 (a) 自然成立.

例 2.2 任给的 $X\in\mathcal{S}^n$, $\theta(X)=\delta_{\mathcal{S}^n_+}(X)$. 那么条件 (a) 等价于 $\operatorname{rank}(\mathcal{A}^*\overline{y}-\overline{\eta})\geqslant n-1$[39,Proposition 3.2].

例 2.3 任给的 $X\in X$, $\theta(X)=\|X\|_{(k)}$. 那么如果以下两个条件中的任一个成立: (i) $\|\mathcal{A}^*\overline{y}-\overline{\eta}\|_*<k$ 且 $\sigma_2(\mathcal{A}^*\overline{y}-\overline{\eta})<1$; (ii) $\|\mathcal{A}^*\overline{y}-\overline{\eta}\|_*=k$, $\sigma_2(\mathcal{A}^*\overline{y}-\overline{\eta})<1$ 并且 $\sigma_n(\mathcal{A}^*\overline{y}-\overline{\eta})>0$, 则集合 $\mathcal{G}_P^1(\overline{y})$ 就是一个多面体集合. 这一结果可以由 $\partial\|\cdot\|_{(k)}$ 的刻画[9,41]中得到.

其次, 性质 2.3 中的条件 (b) 可以被看成广义方程 $0\in -X+\partial\theta(\mathcal{A}^*y-\overline{\eta})$ 在 $(\overline{X},\overline{y})$ 处的严格互补条件. 特别的, 关于 $\theta=\delta_{\mathcal{S}^n_+}$ 的严格互补条件的刻画, 可以在文 [15, Example 4.79] 中找到; 关于 $\theta=\|\cdot\|_{(k)}$ 的严格互补条件的刻画在文 [42] 中找到. 最后, 值得指出的是为了保证定理 2.1 中条件 (c) 成立, 只需存在问题 (2.1) 的一个 KKT 点 $(\widehat{X},\overline{y})$ 满足谱函数 θ 的严格互补条件. 这里的 $\widehat{X}$ 可以与当前考虑的最优解 $\overline{X}$ 不同, 即我们并不要求严格互补条件在最优解 $\overline{X}$ 处成立.

3 增广拉格朗日法的快速收敛率—扰动分析性质的一个应用

下面我们以增广拉格朗日法的快速线性收敛率为例子, 来说明前面讨论的扰动分析性质, 在实际应用特别是优化算法收敛性分析中的重要作用. 以下我们将集中讨论凸矩阵优化模型 (2.1) 及其对偶问题 (2.2).

回顾如前, 我们定义变量 $Z=(y,w,S)$ 对于任给的 $y\in\mathbb{R}^e$, $w\in\mathbb{R}^d$ 和 $S\in\mathbb{X}$, 空间 $\mathbb{Z}=\mathbb{R}^e\times\mathbb{R}^d\times\mathbb{X}$. 令 $c>0$ 是一个正参变量. 对于任给的 $Z\in\mathbb{Z}$ 和 $X\in\mathbb{X}$, 针对对偶问题 (2.2) 的增广拉格朗日函数如下给出

$$L_c(Z;X):=L(Z,X)+(c/2)\|\mathcal{A}^*y+\mathcal{F}^*w+S-C\|^2. \tag{3.1}$$

给定一数量序列 $c_k\uparrow c_\infty\leqslant\infty$ 以及初始点 $X^0\in\mathbb{X}$, 增广拉格朗日法 (ALM) 的第 $(k+1)$

次迭代可以如下给出

$$\begin{cases} Z^{k+1} \approx \arg\min\limits_{Z\in\mathbb{Z}}\{\zeta_k(Z) := L_{c_k}(Z; X^k)\}, \\ X^{k+1} = X^k + c_k(\mathcal{A}^*y^{k+1} + \mathcal{F}^*w^{k+1} + S^{k+1} - C), \end{cases} \quad k \geqslant 0. \tag{3.2}$$

对于凸优化问题, 增广拉格朗日法是非精确的对偶临近点算法的一个特殊应用[43]. 特别的, 令 $P_k := (\mathcal{I} + c_k\mathcal{T}_\phi)^{-1}$, 其中 $\mathcal{I}$ 是 $\mathbb{X}$ 上的单位映射, 而 $\mathcal{T}_\phi$ 为 (2.5) 定义的极大单调的集值映射. 求解优化问题 (2.1) 的临近点算法 (PPA) 可以简单的表示为如下迭代格式

$$X^{k+1} \approx P_k(X^k). \tag{3.3}$$

相应的终止准则为

(A) $\|X^{k+1} - P_k(X^k)\| \leqslant \varepsilon_k, \quad \varepsilon_k \geqslant 0, \quad \sum\limits_{k=0}^{\infty}\varepsilon_k < \infty,$

(B) $\|X^{k+1} - P_k(X^k)\| \leqslant \eta_k\|X^{k+1} - X^k\|, \quad \eta_k \geqslant 0, \quad \sum\limits_{k=0}^{\infty}\eta_k < \infty.$

以下结果揭示了由 ALM(3.2) 生成的求解对偶问题 (2.2) 的迭代序列和 PPA(3.3) 生成的求解原问题 (2.1) 的迭代序列之间的关系. 它的证明可以从文 [42, Proposition 6] 适当改写得来.

性质 3.1 给定 $X^k \in \mathbb{X}$, $Z^{k+1} = (y^{k+1}, w^{k+1}, S^{k+1}) \in \mathbb{Z}$ 和正参数 c_k 对于某个 $k \geqslant 0$. 令 ζ_k 和 P_k 分别如 (3.2) 和 (3.3) 给出. 定义 $X^{k+1} = X^k + c_k(\mathcal{A}^*y^{k+1} + \mathcal{F}^*w^{k+1} + S^{k+1} - C)$. 那么

$$\|X^{k+1} - P_k(X^k)\|^2/(2c_k) \leqslant \zeta_k(Z^{k+1}) - \inf\zeta_k. \tag{3.4}$$

从上述结果, 我们知道如果在求解对偶问题 (2.2) 的 ALM 的乘子序列 $\{X^k\}$ 使用以下两个终止准则 (A′) 和 (B′)

(A′) $\zeta_k(Z^{k+1}) - \inf\zeta_k \leqslant \varepsilon_k^2/2c_k, \quad \varepsilon_k \geqslant 0, \quad \sum\limits_{k=0}^{\infty}\varepsilon_k < \infty,$

(B′) $\zeta_k(Z^{k+1}) - \inf\zeta_k \leqslant (\eta_k^2/2c_k)\|X^{k+1} - X^k\|^2, \quad \eta_k \geqslant 0, \quad \sum\limits_{k=0}^{\infty}\eta_k < \infty,$

那么它可以被视为是使用非精确 PPA 求解原问题 (2.1), 分别使用终止准则 (A′) 和 (B′) 生成的. 因此, 基于以上事实, 我们可以通过研究非精确 PPA 的收敛率来研究 ALM 的收敛率.

下面我们简要回顾一下关于非精确 PPA 收敛率的研究. 在经典文献 [44] 中, Rockafellar 在所谓 $\mathcal{T}_\phi^{-1}$ 在零点处的 Lipschitz 连续假设下(即最优解集映射 $X : \mathbb{X} \rightrightarrows \mathbb{X}$ 在零点处的(鲁棒)孤立平稳性), 得到了非精确 PPA 的收敛率结果. 如前所述, 这一假设自然要求原问题的最优解是唯一的, 这一要求对于一些应用是有局限性的, Luque 将它松弛为一个误差界型的条件[45,(2.1)]. 如果 $\mathcal{T}_\phi$ 是多面体映射, Luque 的条件被证明是成立的[37]. 然而, 当 $\mathcal{T}_\phi$ 是非多面体时, 验证 Luque 的条件可能是困难的, 特别是当最优解集 $X(0) = \mathcal{T}_\phi^{-1}(0)$ 无界时. 在文 [39] 中, 这一条件被进一步放松为最优解集映射 $X : \mathbb{X} \rightrightarrows \mathbb{X}$ 在某个最优点处关于零点的平稳性. 或者集值映射 $\mathcal{T}_\phi$ 在最优点处关于零点的度量次正则性. 基于性质 3.1 和文 [39, Theorem 4.1], 我们可以证明以下增广拉格朗日法求解原问题 (2.1) 的全局和局部(超)线性收敛性.

定理 3.1 假设原问题 (2.1) 的最优解集 $X(0)$ 是非空的. 令 Ψ^* 为对偶问题 (2.2) 的最优值. $\{(Z^k, X^k)\}$ 是 ALM(3.2) 以终止准则 (A$'$) 生成的无限序列, 其中 $Z^k=(y^k,w^k,S^k)$. 那么, 整个序列 $\{X^k\}$ 是有界的并且收敛到某一个最优解 $X^\infty\in X(0)$, 并且序列 $\{Z^k\}$ 满足对所有的 $k\geqslant 0$,

$$\|\mathcal{A}^*y^{k+1}+\mathcal{F}^*w^{k+1}+S^{k+1}-C\|=c_k^{-1}\|X^{k+1}-X^k\|\to 0, \tag{3.5}$$

$$\Psi(Z^{k+1})-\Psi^*\leqslant \zeta_k(Z^{k+1})-\inf\zeta_k+(1/2c_k)(\|X^k\|^2-\|X^{k+1}\|^2). \tag{3.6}$$

进一步, 如果对偶问题 (2.2) 有非空有界解集, 则序列 $\{Z^k\}$ 也是有界的, 并且其任意聚点都是对偶问题 (2.2) 的最优解.

如果原问题 (2.1) 的最优解集映射 $X:\mathcal{X}\rightrightarrows\mathcal{X}$ 在 X^∞ 处关于零点是平稳的(定义), 那么存在常数 $\kappa_p>0$(由问题 (2.1) 决定与算法无关), 在终止准则 (B$'$) 下存在 $\overline{k}\geqslant 0$ 使得对所有 $k\geqslant\overline{k}$,

$$\operatorname{dist}(X^{k+1},X(0))\leqslant\theta_k\operatorname{dist}(X^k,X(0)), \tag{3.7}$$

其中

$$\begin{aligned}&\theta_k=(\mu_k+2\eta_k)(1-\eta_k)^{-1}\quad\text{且}\quad\mu_k=1/\sqrt{1+c_k^2\kappa_p^2},\\&\theta_k\to\theta_\infty=1/\sqrt{1+c_\infty^2\kappa_p^2}\quad(\theta_\infty=0\text{ 如果 }c_\infty=\infty).\end{aligned}$$

此外, 以下不等式揭示了对偶可行性和对偶函数值的 R-(超)线性收敛率:

$$\begin{aligned}&\|\mathcal{A}^*y^{k+1}+\mathcal{F}^*w^{k+1}+S^{k+1}-C\|\leqslant\tau_k^1\operatorname{dist}(X^k,X(0)),\\&\Psi(Z^{k+1})-\Psi^*\leqslant\tau_k^2\operatorname{dist}(X^k,X(0)),\end{aligned}$$

其中

$$\begin{aligned}&\tau_k^1:=c_k^{-1}(1-\eta_k)^{-1}\to\tau_\infty^1=1/c_\infty,\\&\tau_k^2:=\tau_k^1(\eta_k^2\|X^{k+1}-X^k\|+\|X^{k+1}\|+\|X^k\|)/2\to\tau_\infty^2=\|X^\infty\|/c_\infty,\\&(\tau_\infty^1=\tau_\infty^2=0\quad\text{如果}\quad c_\infty=\infty).\end{aligned}$$

定理 3.1 揭示了在原问题 (2.1) 的最优解集映射 $X:\mathbb{X}\rightrightarrows\mathbb{X}$ 在某一个最优解处关于零点的平稳性假设下, 由 ALM 生成的原问题解迭代序列 $\{X^k\}$ 是 Q-(超)线性收敛的, 同时对偶可行性和对偶函数值是至少 R-(超)线性收敛. 更重要的是, 序列 $\{X^k\}$ 的 Q-(超)线性收敛常数 θ_k 可以远小于 1, 即所谓的快速线性收敛率. 例如, 如果参数 c_k 接近 $1/\kappa_p$, 收敛率 θ_k 可以接近 $\sqrt{2}/2$. 这一重要特性使得增广拉格朗日法求解凸优化问题, 特别是大规模矩阵优化问题时, 变得十分有效. 由于篇幅的限制, 我们不在此罗列 ALM 的收敛率在实际应用中的数值验证, 感兴趣的读者可以在许多相关文献中找到, 例如: 求解半定规划问题的 SDPNL[45], SDPNL+[47]; 二次半定规划问题的 QSDPNL[48]; 矩阵核范数极小问题[48]; 矩阵逼近问题[50]; 含 Ky Fan k-范数的矩阵优化问题[2]等. 最后, 我们还需要指出的是虽然定理 3.1 看似只讨论了基于最优解集映射平稳性(即节 2 讨论的结果)的对偶 ALM 的快速收敛率, 然而类似的基于 KKT 系统解映射 $S_{\rm KKT}$ 的(鲁棒)孤立平稳性的所谓原始对偶 PPA 的快速收敛率可以同理得到, 具体结果参见 Rockafellar 的经典文献[43, 44].

4 结论与进一步研究的思考

由于近年来实际问题特别是大数据应用的发展, 矩阵优化问题越来越得到优化研究者, 甚至是其他领域的研究者的高度关注, 成为热点问题. 优化问题的扰动性分析是优化理论研究的基础与核心, 为包括算法设计在内的优化研究提供重要的理论基础. 由于矩阵优化问题的非多面体性, 使得相应扰动分析理论的研究本质上与经典的多面体优化问题(非线性规划)不同. 本文结合文献 [1, 2], 简要介绍了矩阵优化扰动性分析方面取得的若干最新进展: 一般(非凸)含非多面体约束矩阵优化问题的 KKT 系统解映射在最优解处的鲁棒孤立平稳性的等价刻画以及一般凸矩阵优化问题的最优解集映射的平稳性的充分性条件. 最后, 我们通过对增广拉格朗日法在求解凸矩阵优化问题快速收敛率的分析, 展示了矩阵优化扰动性分析理论对算法设计与求解实际问题重要性.

通过本文的简要介绍, 我们看到矩阵优化的相关研究还处于起步阶段, 无论是理论、算法还是实际应用方面还有许多问题没有很好的解决. 即使在扰动性分析方面, 其他一些类 Lipschitz 性质, 如 Aubin 性质都是可以进一步研究的课题. 在算法设计上, 如何进一步利用扰动分析结果来设计快速求解大数据优化问题的有效算法一直有挑战性的问题, 如何将扰动性分析理论与统计优化的相关统计理论有机结合, 以解决大数据应用的理论与算法的重要问题也将逐渐成为热点.

致谢 作者感谢两名匿名审稿人对提高本文的质量和可读性提出的许多宝贵建设性意见. 最后, 作者还要感谢大连理工大学的王石玮同学仔细阅读了本文的初稿, 并对修改稿提出了许多宝贵意见.

参 考 文 献

[1] Ding C, Sun D F, Zhang L W. Characterization of the robust isolated calmness for a class of conic programming problems [J]. *SIAM Journal on Optimization*, 2017, **27**: 67-90.

[2] Cui Y, Ding C, Zhao X Y. Quadratic growth conditions for convex matrix optimization problems associated with spectral functions [J]. *SIAM Journal on Optimization*, 2017, **27**: 2332-2355.

[3] Todd M J. Semidefinite optimization [J]. *Acta Numerica*, 2001, **10**: 515-560.

[4] Vandenberghe L, Boyd S. Semidefinite Programming [J]. *SIAM Review*, 1996, **38**: 49-95.

[5] Ben-Tal A, Nemirovski A. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* [M]. Philadelphia: SIAM, 2001.

[6] Candès E J, Recht B. Exact matrix completion via convex optimization [J]. *Foundations of Computational Mathematics*, 2008, **9**: 717-772.

[7] Candès E J, Tao T. The power of convex relaxation: near-optimal matrix completion [J]. *IEEE Transactions on Information Theory*, 2009, **56**: 2053-2080.

[8] Recht B, Fazel M, Parrilo P A. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization [J]. *SIAM Review*, 2010, **52**: 471-501.

[9] Watson G A. On matrix approximation problems with Ky Fan k-norms [J]. *Numerical Algorithms*, 1993, **5**: 263-272.

[10] Greenbaum A, Trefethen L N. GMRES/CR and Arnoldi/Lanczos as matrix approximation problems [J]. *SIAM Journal on Scientific Computing*, 1994, **15**: 359-368.

[11] Toh K C, Trefethen L N. The Chebyshev polynomials of a matrix [J]. *SIAM Journal on Matrix Analysis and Applications*, 1998, **20**: 400-419.

[12] Boyd S, Diaconis P, Sun J, Xiao L. Fastest mixing Markov chain on a path [J]. *The American Mathematical Monthly*, 2006, **113**: 70-74.

[13] Boyd S, Diaconis P, Parrilo P A, Xiao L. Fastest mixing Markov chain on graphs with symmetries [J]. *SIAM Journal on Optimization*, 2009, **20**: 792-819.

[14] Bonnans J F, Shapiro A. Optimization problems with perturbations: A guided tour [J]. *SIAM review*, 1998, **40**: 228-264.

[15] Bonnans J F, Shapiro A. *Perturbation Analysis of Optimization Problems* [M]. New York: Springer, 2000.

[16] Dontchev A L, Rockafellar R T. *Implicit Functions and Solution Mappings* [M]. New York: Springer, 2009.

[17] Facchinei F, Pang J S. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I* [M]. New York: Springer-Verlag, 2003.

[18] Klatte D, Kummer B. *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications* [M]. Dordrecht: Kluwer, 2002.

[19] Mordukhovich B S. *Variational Analysis and Generalized Differentiation I & II* [M]. Berlin Heidelberg: Springer-Verlag, 2006.

[20] Rockafellar R T, Wets R J B. *Variational Analysis* [M]. Berlin: Springer, 1998.

[21] Rockafellar R T. *Convex Analysis* [M]. Princeton: Princeton University Press, 1970.

[22] Zowe J, Kurcyusz S. Regularity and stability for the mathematical programming problem in Banach spaces [J]. *Applied Mathematics and Optimization*, 1979, **5**: 49-62.

[23] Dontchev A L, Rockafellar R T. *Characterizations of Lipschitzian stability in nonlinear programming* [M]//*Mathematical Programming With Data Perturbations*, New York: Marcel Dekker, 1997, 65-82.

[24] Levy A B. Implicit multifunction theorems for the sensitivity analysis of variational conditions [J]. *Mathematical Programming*, 1996, **74**: 333-350.

[25] Mordukhovich B S, Outrata J V, Ramŕez H. Graphical derivatives and stability analysis for parameterized equilibria with conic constraints [J]. *Set-Valued and Variational Analysis*, 2015, **23**: 687-704.

[26] Robinson S M. Generalized equations and their solutions, Part II: Applications to nonlinear programming [J]. *Mathematical Programming Study*, 1982, **19**: 200-221.

[27] Klatte D. Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs [J]. *Mathematical Programming Series B*, 2000, **88**: 285-311.

[28] Kyparisis J. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming [J]. *Mathematical Programming*, 1985, **32**: 242-246.

[29] Bonnans J F, Cominetti R, Shapiro A. Sensitivity analysis of optimization problems under second order regular constraints [J]. *Mathematics of Operations Research*, 1998, **23**: 806-831.

[30] Mangasarian O L. A simple characterization of solution sets of convex programs [J]. *Operations Research Letters*, 1988, **7**: 21-26.

[31] Zhou Z R, So A M C. A unified approach to error bounds for structured convex optimization problems [J]. *Mathematical Programming*, 2017, **165**: 689-728.

[32] Bauschke H H, Borwein J M. On projection algorithms for solving convex feasibility problems [J]. *SIAM Review*, 1996, **38**: 367-426.

[33] Bauschke H H, Borwein J M, Li W. Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization [J]. *Mathematical Programming*, 1999, **86**: 135-160.

[34] Lewis A S. The convex analysis of unitarily invariant matrix functions [J]. *Journal of Convex Analysis*, 1995, **2**: 173-183.

[35] Lewis A S. Convex analysis on the Hermitian matrices [J]. *SIAM Journal on Optimization*, 1996, **6**: 164-177.

[36] Sun J. *On Monotropic Piecewise Qudratic Programming* [D]. Seattle: University of Washington, 1986.

[37] Robinson S M. *Some continuity properties of polyhedral multifunctions* [M]// *Mathematical Programming at Oberwolfach*, Berlin: Springer, 1981, 206-214.

[38] Cui Y. *Large scale composite optimization problems with coupled objective functions: theory, algorithms and applications* [D] Singapore: National University of Singapore, 2016.

[39] Cui Y, Sun D F, Toh K C. On the asymptotic superlinear convergence of the augmented Lagrangian method for semidefinite programming with multiple solutions [J]. arXiv: 1610.00875, 2016.

[40] Tseng P. Approximation accuracy, gradient methods, and error bound for structured convex optimization [J]. *Mathematical Programming*, 2010, **125**: 263-295.

[41] Overton M, Womersley R. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices [J]. *Mathematical Programming*, 1993, **62**: 321-357.

[42] Ding C. Variational analysis of the Ky Fan k-norm [J]. *Set-Valued and Variational Analysis*, 2017, **25**: 265-296.

[43] Rockafellar R T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming [J]. *Mathematics of Operations Research*, 1976, **1**: 97-116.

[44] Rockafellar R T. Monotone operators and the proximal point algorithm [J]. *SIAM Journal on Control and Optimization*, 1976, **14**: 877-898.

[45] Luque F J. Asymptotic convergence analysis of the proximal point algorithm [J]. *SIAM Journal on Control and Optimization*, 1982, **22**: 277-293.

[46] Zhao X Y, Sun D F, Toh K C. A Newton-CG augmented Lagrangian method for semidefinite programming [J]. *SIAM Journal on Optimization*, 2010, **20**: 1737-1765.

[47] Yang L Q, Sun D F, Toh K C. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints [J]. *Mathematical Programming Computation*, 2015, **7**: 331-366.

[48] Li X D, Sun D F, Toh K C. QSDPNAL: A two-phase proximal augmented Lagrangian method for convex quadratic semidefinite programming [J]. arXiv:1512.08872, 2015, 1-35.

[49] Liu Y J, Sun D F, Toh K C. An implementable proximal point algorithmic framework for nuclear norm minimization [J]. *Mathematical Programming*, 2011, **133**: 399-436.

[50] Chen C, Liu Y J, Sun D F, Toh K C. A semismooth Newton-CG based dual PPA for matrix spectral norm approximation problems [J]. *Mathematical Programming*, 2014, **155**: 435-470.